

# The STREAMES Project: Linking Heuristic And Empirical Knowledge Into An Expert System To Assess Stream Managers

**J. Comas<sup>1</sup>, E. Llorens<sup>1</sup>, M. Poch<sup>1</sup>, G. Markakis<sup>2</sup>, T. Battin<sup>3</sup>, S. Gafny<sup>4</sup>, E. Maneux<sup>10</sup>, E. Marti<sup>5</sup>, M. Morais<sup>6</sup>,  
M.A. Puig<sup>5</sup>, M. Pusch<sup>7</sup>, J.L. Riera<sup>8</sup>, F. Sabater<sup>8</sup>, A.G. Solimini<sup>9</sup> and P. Vervier<sup>10</sup>**

<sup>1</sup> Chemical and Environmental Engineering Laboratory (LEQUIA), University of Girona, Campus Montilivi s/n, E-17071 Girona, Catalonia, EU, e-mail: {quim/esther}@lequia.udg.es

<sup>2</sup> Natural History Museum of Crete, University of Crete, Knossou Avenue, 714 09 Heraklion, Greece, EU

<sup>3</sup> Department of Limnology, Institute of Ecology and Conservation Biology, Althanstrasse 14, 1090 Viena, Austria, EU

<sup>4</sup> Institute for Nature Conservation Research, Ramat Aviv, 699778 Tel Aviv, Israel

<sup>5</sup> Centre d'Estudis Avançats de Blanes (CSIC), Camí de Sta. Bàrbara s/n, 17300 Blanes, Catalonia, EU

<sup>6</sup> Centro de Ecologia Aplicada, Universidad de Evora, Largo dos Colegiais 2, 7002-554 Evora, Portugal, EU

<sup>7</sup> Institut fuer Gewaesseroekologie und Binnenfisherei, Mueggelseedamm, 310, 12587 Berlin, Germany, EU

<sup>8</sup> Departament d'Ecologia, Universitat de Barcelona, Diagonal 645, Barcelona 08028, Catalonia, EU

<sup>9</sup> Dipartimento di Biologia, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Roma, Italy, EU

<sup>10</sup> Centre d'Ecologie des Systemes Aquatiques Continentaux – UMR C5576 – Centre National de la Recherche Scientifique - Université Paul Sabatier, Rue Jeanne Marvig 29, 31055 Toulouse, France, EU

**Abstract:** The increase in stream nutrient loads from anthropogenic sources has become a serious problem, especially in developed regions. Humans affect streams by modifying the landscape in ways that increase the transport of nutrients to surface waters, by directly dumping urban or industrial sewage into the stream, or by modifying streams in ways that reduce their ability to respond to increased nutrient loads. In Mediterranean regions these problems are compounded by the scarcity of water. The decision-making processes involved in water quality management require extensive human expertise or extensive computation with large data sets. In this sense, the STREAMES project aims to develop a knowledge-based environmental decision support system (EDSS) to support and advice water managers in the management of human-altered streams. This EDSS will integrate an Expert System (ES), concretely a rule-based reasoning system (RBS), with a Geographical Information System to address spatial information for the appropriate stream management actions, and a numerical model to estimate point and non-point nutrient sources from middle size catchments. The RBS will be developed by integrating heuristic knowledge from experts in surface water management, as well as empirical knowledge from stream scientists, based both on previous studies and on data directly acquired from experimental sampling. This paper will present the objectives of the STREAMES project with emphasis in the knowledge acquisition and development of the RBS.

**Keywords:** expert system; rule-based system, environmental decision support system; stream management; river water quality

## 1. INTRODUCTION

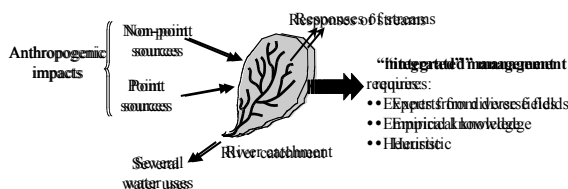
Nowadays, poor river water quality has become a serious problem, especially in developed regions, due to the high nutrient (nitrogen and phosphorus) loads from anthropogenic sources dumped into the rivers.

Humans affect streams:

- By modifying the land uses in ways that increase the transport of nutrients to surface waters (large inputs of nutrients from non-point sources, i.e. agricultural activity),
- by directly dumping inputs from point sources (urban or industrial sewage),
- by modifying the streams in ways that reduce their ability to respond to increased nutrient loads (e.g. elimination of meanders, destruction of the riparian vegetation, etc...).

Pristine stream ecosystems can cope with a certain degree of pollution, whereas heavily polluted and modified streams cannot retain and transform excessive nutrient loads. Nevertheless, the response of streams to anthropogenic impacts is not invariable but different according to the type of river and water uses. In Mediterranean regions these problems are compounded by the scarcity of water.

Therefore, the decision-making processes involved in stream reach management require extensive human expertise from people involved directly with day-to-day stream problems (water managers), empirical knowledge from scientific research and elaborated calculation over large sets of numerical and symbolic data.



**Figure 1.** The management problem

Thus, the stream water quality management becomes a typical complex and ill-defined problem whose optimal management requires an integrated and multidisciplinary approach (figure 1). This integrated management can be reached with a tool built upon the concepts and methods of human reasoning, an intelligent tool. In this sense, the STREAMES

(STream REach Management, an Expert System<sup>1</sup>) project appears as an attempt to develop and implement a knowledge-based environmental decision support system (KB-EDSS or simply, EDSS) to help Water Managers (WM) in taking decisions.

This paper describes the objectives of STREAMES with major emphasis in the bottleneck of the RBS development: the knowledge acquisition step.

The organisation of this paper follows the next sections: Section 1 introduces the problem of nutrient retention. Section 2 gives a brief description of the Streames project. The EDSS characteristics are introduced in Section 3. Section 4 addresses the issue of the knowledge acquisition as a key step for developing a user-friendly tool. Section 5 presents the future work with respect to the EDSS and finally, some conclusions are given in Section 6.

## 2. THE STREAMES PROJECT

The application of Artificial Intelligence in environmental sciences is a relatively new discipline. Papers began to be published sporadically in the middle of 80s, but it was not until the 90s when the number of papers experienced a significant increase [Cortés et al., 2000]. Some interesting environmental applications include MEDEX [Hadjimichael et al., 1996], an intelligent tool to assist Mediterranean weather forecast; CHARADE [Avesani et al., 1993], a decision-making system for environmental emergencies; FRAME [Calori et al., 1994], an ES designed to aid in the selection of air pollution models; and DAI-DEPUR, a distributed knowledge-based system to supervise wastewater treatment plant management ([Sánchez et al., 1996] and [R.-Roda et al., 2002] for a recent vision of the application). Within the river water domain, mathematical models, simulation models and decision support systems are used for catchment management, to reduce nutrient loads or to solve eutrophication problems ([Rekolainen et al., 1999], [Young et al., 1995] and [Davis et al., 1998]). However, none of these tools implies cooperation between stream managers and scientists. Moreover, they only consider descriptive parameters. A new approach integrating expertise, existing knowledge and new empirical knowledge about

<sup>1</sup> ("Human effects on nutrient cycling in fluvial ecosystems: The development of an ES to assess stream water quality management at reach scale", EVK1-2000-00081, Vth Framework Programme EC; [www.streames.org](http://www.streames.org))

nutrient retention capacity concerning not only descriptive and structural but also functional parameters is proposed. In this context, STREAMES aims at evaluating the effect of substantial nutrient loads on in-stream retention. Furthermore, STREAMES will examine the relationships between nutrient retention and selected physical, chemical and biological structural or functional parameters that may constrain (i.e., nutrient sources from the catchment) or control (i.e., in-stream processes) nutrient retention capacity in human altered streams. Particular emphasis will be on Mediterranean streames.

In a practical sense, the final goal of STREAMES project is to develop and validate an EDSS for stream managers from either private or public water quality agencies. So the end product will be the EDSS, a useful decision support system to give diagnosis about the stream quality and propose solutions to the problems. These solutions should allow to attain not only a water quality improvement but also a good ecological state for the river, according to the Water Framework Directive (2000/60/EEC).

This project involves 17 partners (10 from scientific research centres or universities and 7 from water agencies) from 8 European countries and the project is divided into five workpackages (WP). Workpackage 1 analyses, at the catchment scale, the relationship between land uses and nutrient loads to ecosystems and the relative importance of point and non-point nutrient sources. WP2 analyses the effects of high nutrient loads on stream nutrient transport, transformation and retention at reach scale. WP3 analyses the role of stream biota on the control of nutrient retention at the sub-reach scale. Finally, WP4 focuses on the development of the EDSS and WP5 promotes the utility and application of the EDSS to other end-users.

### 3. THE ENVIRONMENTAL DECISION SUPPORT SYSTEM

#### 3.1. Structure of the EDSS

The EDSS to be developed in the STREAMES project integrates a rule-based reasoning component (the core of the system) with a Geographic Information System (GIS) and linked with a numerical model to estimate point and non-point sources. This model will be the MONERIS model [Behrendt et al., 2001], modified for Mediterranean regions.

The EDSS must be fed with all the available information relevant to the final outcome. This information, both qualitative and quantitative, is stored in a database, where it remains easily accessible to software requirement. Additional information can be required to the user to reach a deeper conclusion. Then the user would be responsible for gathering these new measurements. The amount and quality of this information will condition the quality of the EDSS's outcomes.

The EDSS will provide 3 types of outputs: diagnosis, actions and prognosis. Figure 2 shows the conceptual framework of the EDSS with its three-level outcome, based on the specific requirements of the WM. Concerning the diagnosis step, the EDSS will be able to infer the river state related to functionality features (for example, the self-purification capacity of the stream reach relative to its potential capacity, the nutrient uptake length or the recovery time). Once this diagnosis step is reached, the user can require the EDSS to propose a list of suitable management strategies to maximise self-purification at the stream reach.

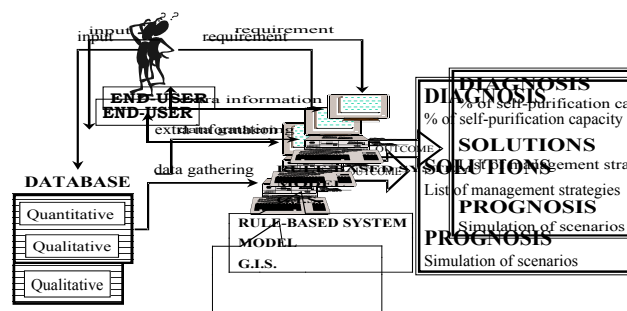


Figure 2. How the EDSS will support the decision making tasks on stream management.

Finally, the EDSS will offer a prognosis output to the end-user, providing several scenarios to simulate the effect of the different actions proposed as solutions and giving the percentage of success according to the adopted solution.

#### 3.2. Structure of the Rule-Based System

The structure of any RBS presents two main independent modules: the Knowledge Base (KB) and the Inference Engine (IE). The KB contains the overall knowledge of the process (in our case, stream nutrient management) codified by means of heuristic rules (a rule is a set of conditions and conclusions linked to a

given hypothesis). The bottleneck of the KB development is the knowledge acquisition process. In the first half of the project, G2 [Gensym, 2000] will be used as inference engine to build the first prototype of the RBS. However, the final product must be codified over a new shell (including the inference engine) that must also be developed during the course of the project by WP4.

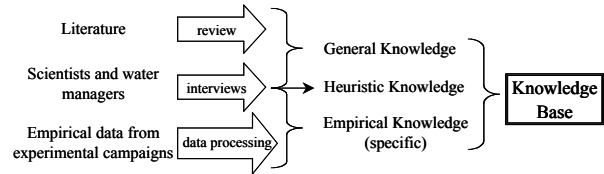
### 3.3. Requirements of the EDSS

The EDSS will support WM in two ways: First, it will help them to evaluate the sources and magnitude of nutrient (nitrogen and phosphorus) loads affecting the stream reach of interest. Second, it will help WM to decide on the best strategy for stream amelioration at the particular reach, with special emphasis on actions directed towards increasing nutrient retention and transformation within the stream.

The EDSS must detect any problems caused for a wastewater treatment plant (WWTP). It must predict which site is the best place to build a new WWTP and which wastewater treatment has to be implemented. Moreover, the EDSS will evaluate the WWTP impact on the studied reach and give a set of actions to subdue the impacts caused by the WWTP or other inputs. In order to prevent confusions, the EDSS will only focus on the reach scale. Of course, in some problems, the EDSS must work on the catchment scale.

## 4. KNOWLEDGE ACQUISITION

The knowledge acquisition process is the key process to build a complete knowledge base, in this case, a manual of operation on stream management, in the development of the rule-based system, the main module of the EDSS. We are interested on acquiring the whole knowledge necessary to identify the stream quality problems and to solve them. This knowledge is not only concerned on how the WM deal with their rivers but also on how the scientists think that stream management could be improved. Three types of knowledge can be distinguished: general knowledge related to the domain, heuristic knowledge and empirical knowledge, more specific since contains distinguishing features related to the specific sites where the RBS will be applied (Figure 3).



**Figure 3.** Knowledge acquisition in the RBS development.

In order to obtain as much as possible of knowledge of quality about nutrient retention, different sources will be used:

- literature
- survey among stream experts and face-to-face interviews
- study of the historical data and sampling sites of eleven different scenarios

While the general and heuristic knowledge of the process can be mainly extracted from literature and human experts, the empirical knowledge can only be acquired by processing empirical data directly collected from the experimental campaigns in each study site (in our case from WP1, WP2 and WP3 tasks) (figure 3).

### 4.1. Literature review

A bibliographical review about stream reach management, especially referred to nutrient pollution, and the state-of-the-art of decision support systems applied to river or stream management will be done.

### 4.2. Questionnaire and face-to-face interviews

A questionnaire will be sent to the WM and scientists. The purpose of this questionnaire is to obtain a general view of stream management. The objective is to acquire as much information as possible to understand the reasoning mechanisms of WM (and also from scientists) when tackling any problem about stream nutrient retention. Therefore, information about data availability, minimal specifications, expectations and minimal capabilities of the RBS are required. With the results of this survey, we will prepare the questions for the personal interviews.

Each questionnaire consists of 4 parts:

- Part 1 intends to obtain general information about river basins of each participant country.
- Part 2 refers to the stream water quality. Availability and format of historical data and relevance of most important parameters to identify the stream water quality are requested.
- Part 3 relates to the main stream quality problems that WM have to face when managing rivers and the criteria used to identify them. This part intends to identify the minimum data and knowledge used for WM (minimum required data for the RBS) to tackle any problem.
- Part 4 questions about requirements and capabilities of the whole EDSS. It asks about the preferences of WM with respect to the possible outputs of the EDSS: diagnosis, range of solutions and types of prognosis.

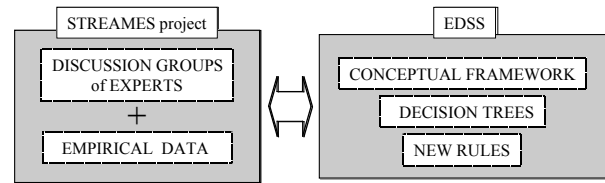
#### 4.3. Empirical data from sampling sites

The behaviour of pristine rivers and polluted rivers will be studied within the experimental campaigns (eleven scenarios with different conditions are considered). Both types of rivers can react differently against high nutrient loading. We expect to find some relationships between functional and descriptive parameters from the processing and study of data obtained from sampling sites. Some inductive machine learning tools can be used to improve knowledge discovering from empirical data ([Comas et al., 2001] and [R.-Roda et al., 2001]).

#### 5. FUTURE STEPS

Once all knowledge will be gained (empirical from experimental campaigns, questionnaires and internal forums of discussion and general from questionnaires and literature research), efforts will concentrate on (see figure 4):

- redefine the conceptual framework of the EDSS according to the WM (end-users) final specific requirements in order to build a useful support tool,



**Figure 4.** Translating knowledge into a useful EDSS.

- structuring and representing the knowledge in a decision tree (DT) fashion as a previous step to build the KB. Afterwards, each branch of the trees will be easily codified by means of a set of heuristic “if-then” rules. We are also evaluating the use of some automatic tools to translate the DT into rules.

Every decision tree will refer to a defined problem: eutrophication, excess of ammonia, excess of phosphorus, suspended solids, organic matter pollution...but probably most of them will be interrelated to others.

- After the development of the KB, the RBS will be integrated with spatial information within the context of a GIS system. The final EDSS must be able to communicate with different external applications (e.g. an external database).
- Finally, a global validation of the prototype will be done. This validation will be conducted at two different levels: (a) a validation of the technical aspects of the prototype in relation to the use of this tool by final end-users; and (b), a validation of the outputs of the EDSS.

#### 6. CONCLUSIONS

This paper presents a European project whose main deliverable is an EDSS prototype to support and advice water agencies in the management of human-altered streams. The complex management of environmental systems requests this kind of tools to be applied. This project presents an important example of application of science into the real world.

As far as we know, this will be the first time that an EDSS is proposed to estimate and evaluate uncertainties in the prediction of stream water quality and to support the decision of management actions. Another important issue is the integration of a Geographic Information System to address spatial information for the appropriate stream management actions.

Additionally, other innovative aspects of this project are a) the conduction of field research that will support the

development of the RBS, b) the close involvement of WM/authorities throughout the planning, implementation and validation of this tool, c) the tight collaboration that will exist among scientist, managers and RBS developers during the course of the project. This collaboration will ensure the consideration of the stream water quality problem from several perspectives and, therefore, the resultant product will benefit from it. We are aware that this may be a challenge, but at the same time we strongly believe that it is only through a close dialogue between academic/research and management institutions that stream management strategies can be developed and applied successfully.

## 7. REFERENCES

- Avesani, P., A. Perini, and F. Ricci, Combining CBR and constraint reasoning in planning forest fire fighting. *Proceedings of the First European Workshop on Case-Based Reasoning*, 235-239, 1993.
- Behrendt, H., P. Huber, M. Ley, D. Opitz, O. Schmoll, G. Scholz, and R. Uebe, *Nutrient emissions into river basins of Germany*. UBA-Text and CD-ROM, 75/99, 288 S, 1999.
- Calori, G., F. Colombo, and G. Finzi, Frame: a knowledge-based tool to support the choice of the right air pollution model, *Computer Support for Environmental Impact Assessment*, Guariso G. and B. Page (Eds.), 211-222, 1994.
- Comas, J., Dzeroski, S., Gibert, K., R.-Roda, I. and Sánchez, M., Knowledge Discovery by means of Inductive Methods in Wastewater Treatment Plant Data, *AI communications*, 14(1), 45-62, 2001.
- Cortés, U., M. Sánchez-Marrè, L. Cecaronni, I. R.-Roda, and M. Poch, Environmental Decision Support Systems, *Applied Intelligence* 13(1), 77-91, 2000.
- Davis, J. R., T.F.N. Farley, W.J. Young, and S.M. Cuddy, The experience of using a decision support system for nutrient management in Australia, *Water Science and Technology*, 37(3), 209-216, 1998.
- Gensym. G2 Classic 5.1 Rev.7, Gensym Corporation, Cambridge, USA, 2000.
- Hadjimichael, M., P. Arunas, L. Kuciauskas, and R. Brody, MEDEX: a fuzzy system for forecasting Mediterranean gale force winds. *Proceedings FUZZ-IEEE96, International Conference on Fuzzy systems*, 529-534, 1996.
- Rekolainen, S., J. Grönroos, I. Bärlund, A. Nikander, and Y. Laine, Modelling the impacts of management practices on agricultural phosphorus losses to surface waters of Finland, *Water Science and Technology*, 39(12), 265-272, 1999.
- R.-Roda, I., Comas, J., Poch, M., Sánchez-Marrè, M. and Cortés, U., Automatic Knowledge Acquisition from Complex Processes for the development of Knowledge Based Systems, *Industrial and Engineering Chemistry Research*, 40(15), 3353-3360, 2001.
- R.-Roda, I., Comas, J., Colprim, J., Poch, M., Sánchez-Marrè, M., Cortés, U., Baeza, J. and Lafuente J., A hybrid supervisory system to support wastewater treatment plant operation: implementation and validation, *Water Science & Technology*, 45(4-5), 289-297, 2002.
- Sánchez, M., U. Cortés, J. Lafuente, I. R.-Roda, and M. Poch., DAI-DEPUR: a Distributed Architecture for Wastewater Treatment Plants Supervision. *Artificial Intelligence in Engineering*, 10(3), 275-285, 1996.
- Young, W.J., T.F. Farley, and J.R. Davis, Nutrient management at the catchment scale using a decision support system, *Water Science and Technology* (River basin management for sustainable development), 32(5-6), 277-282, 1995.